

Analisis Properti Psikometri Subtes Merkaufgaben (ME) dengan Rasch Model

Aprili Wahyu Hakiki, Ahyani Radhiani Fitri, Ivan Muhammad Agung

Fakultas Psikologi Universitas Islam Negeri Sultan Syarif Kasim Riau
email: hakiki.apriilihakiki@gmail.com

Abstrak

Merkaufgaben (ME) merupakan subtes Intelligenz Struktur Test (IST) yang mengukur kemampuan mengingat seseorang. Penelitian ini bertujuan untuk mengevaluasi karakteristik psikometri subtes ME IST 70 yang kemudian dilanjutkan dengan merevisi aitem yang memiliki karakteristik psikometri yang buruk. Penelitian ini dibagi kedalam dua tahapan: Pertama, mengestimasi properti psikometris subtes Merkaufgaben (ME) dengan menggunakan Rasch Model (N=293), hasilnya menunjukkan bahwa terdapat 1 aitem misfit (aitem nomor 167) dan 2 aitem yang terjangkit DIF (aitem nomor 160 dan 166). Kedua, melakukan uji coba subtes ME yang telah direvisi (N=124), hasilnya menunjukkan bahwa aitem yang direvisi memenuhi kriteria aitem fit dan tidak terjangkit DIF ($p=0.0670-1.000$, $p>0.05$) dengan koefisien reliabilitas responden ($\alpha=0.76$), aitem ($\alpha=0.97$), serta instrumen ($\alpha=0.77$) yang termasuk pada kategori bagus.

Kata kunci: *Intelligenz Struktur Test (IST), Merkaufgaben (ME), model rasch*

Analysis of Psychometric Properties of Merkaufgaben (ME) Subtest Using Rasch Model

Abstract

Merkaufgaben (ME) is Intelligenz Struktur Test (IST) subtest which measures person's ability to memorize. The purpose of this study was to evaluate psychometric characteristics of ME IST 70 subtest and to continue by revising the items that had bad psychometric characteristics. This study was divided into two steps: First, estimated psychometric properties of Merkaufgaben (ME) subtest using rasch model (N=293), the result showed there were one misfit item (item no.167) and 2 items detected by DIF (items no.160 and 166). Second, performed Merkaufgaben (ME) subtest that had been revised (N=124), which showed the items that had been revised was fit items criteria and not detected by DIF ($p=0.0670-1.000$, $p>0.05$) with respondent reliability coefficient ($\alpha=0.76$), the items ($\alpha=0.97$), and the instruments ($\alpha=0.77$) were in good category.

Keywords: *Intelligenz Struktur Test (IST), Merkaufgaben (ME), rasch model*

Pendahuluan

Ingatan memiliki peranan penting dalam proses belajar. Hal ini berkaitan dengan fungsi mengingat sebagai proses menyimpan hal-hal yang sudah diketahui untuk dikeluarkan kembali pada saat yang lain (Sarwono, 2010). Mengingat merupakan salah satu tolak ukur dalam pengukuran kecerdasan seseorang. Beberapa alat tes inteligensi menggunakan aspek mengingat dalam pengukuran kecerdasan, seperti: Binet

yang memiliki aspek mengingat pada subtes penalaran kuantitatif (Sobur, 2003); Weschler pada skala verbal (Azwar, 2011); dan tes IST (Intelligenz Struktur Tes) pada salah satu subtesnya, yaitu Merkaufgaben (ME).

IST merupakan alat tes yang sering digunakan baik di lingkungan pendidikan maupun pekerjaan meskipun usianya sudah lebih dari 40 tahun (Rahmawati, 2014; Bawono, 2008). IST (Intelligenz Struktur Test) dikembangkan oleh Rudolf Amthauer di Frankfurt, Jerman pada tahun 1953. Salah

satu subtes pada IST yang mengukur aspek mengingat ialah ME (Merkaufgaben). ME berasal dari bahasa Jerman, yaitu Merkaufgaben yang berarti ingat-tugas (Adiwimarta, Darmojuwono, & Hastrick, 2011).

Beberapa penelitian mengenai uji karakteristik psikometri terhadap subtes ME pada IST telah dilakukan dan menunjukkan hasil yang berbeda. Penelitian yang dilakukan oleh Bawono (2008) menjelaskan ME valid dan cukup reliabel dalam mengukur Long-Term Memory (LTM) seseorang.

Penelitian yang dilakukan oleh Rahmawati (2014) menunjukkan bahwa terdapat 6 dari 20 aitem subtes ME memiliki karakteristik psikometri yang tidak baik dan perlu direvisi. Berdasarkan analisis tersebut, maka aitem yang tergolong baik ialah nomor 157, 158, 159, 161, 166, dan 172 sedangkan aitem yang tergolong tidak baik ialah nomor 160, 162, 163, 164, 165, 167, 168, 169, 170, 171, 173, 174, 175, dan 176.

Hasil penelitian yang dilakukan oleh Agung dan Fitri (2016) pada mahasiswa UIN Suska Riau menunjukkan bahwa subjek lebih memiliki kecenderungan untuk menjawab aitem dengan kategori tidak baik, sehingga aitem-aitem yang tidak baik dalam batasan psikometri perlu dilakukan revisi.

ME merupakan subtes yang mengukur aspek mengingat dan hal penting dalam pengukuran kecerdasan, khususnya IST. Mengingat beberapa aitem pada subtes ME memiliki karakteristik psikometri yang tidak baik dan perlu direvisi, maka apakah masih tepat dan dipercaya dalam pengukuran kemampuan mengingat (Rahmawati, 2014; Bawono, 2008, Agung & Fitri, 2016). Oleh karena itu, peneliti tertarik untuk melakukan evaluasi karakteristik psikometri subtes ME dan melakukan revisi terhadap aitem-aitem yang tergolong tidak baik dalam batasan psikometri menggunakan item response theory.

Penelitian ini bertujuan untuk mengevaluasi subtes ME (Merkaufgaben) apakah masih memiliki karakteristik psikometri yang baik dan merevisi aitem pada subtes ME. Aitem-aitem yang tergolong kategori tidak baik sesuai batasan psikometri akan dilakukan revisi dan akan dilihat perbedaan nilai

koefisien dari karakteristik psikometri soal subtes ME (Merkaufgaben) pada IST sebelum dan sesudah revisi.

Karakteristik pengukuran psikometri adalah perbandingan antara atribut yang diukur dengan alat ukurnya, hasilnya dinyatakan secara kuantitatif, dan bersifat deskriptif (Azwar, 2009). Pada penelitian ini, karakteristik psikometri didasarkan pada pendekatan item response theory.

Model paling sederhana dalam item response theory adalah model logistik 1 parameter yang dikenal sebagai model Rasch yang menggunakan parameter kesulitan aitem (b) untuk membedakan antar aitem (Embretson, 2000). Pemodelan Rasch mengkomodasi pendekatan probabilitas dalam memandang atribut sebuah objek ukur yang bersifat deterministik, skor murni yang bebas dari error pengukuran, data interval, tahan terhadap data hilang, dan pengukuran secara objektif (Sumintono & Widhiarso, 2015).

Karakteristik psikometri menggunakan pendekatan item response theory terdiri dari beberapa hal, yaitu:

Reliabilitas

Reliabilitas menjelaskan seberapa jauh pengukuran yang dilakukan berkali-kali akan menghasilkan informasi yang sama (Sumintono & Widhiarso, 2013). Reliabilitas dalam pemodelan Rasch didasarkan pada konsistensi internal yang menekankan pada konsistensi butir-butir yang terdapat dalam instrumen. Reliabilitas dilihat dari hasil summary statistic melalui koefisien cronbach alpha KR-20. Kriteria nilai yang digunakan ialah < 0.5 (buruk); 0.5 – 0.6 (jelek); 0.6 – 0.7 (cukup); 0.7 – 0.8 (bagus); dan > 0.8 (bagus sekali) (Sumintono & Widhiarso, 2015).

Unidimensionalitas

Unidimensionalitas adalah ukuran yang penting untuk mengevaluasi apakah instrumen yang dikembangkan mampu mengukur apa yang seharusnya diukur (Sumintono & Widhiarso, 2013). Independensi lokal juga berhubungan dengan jumlah variabel trait yang mendasari aitem. Pada pemodelan Rasch, analisis unidimensionalitas meng-

gunakan analisis komponen utama (Principal Component Analysis) dari residual, yaitu mengukur sejauh mana keragaman dari instrumen mengukur apa yang seharusnya diukur (Sumintono & Widhiarso, 2015). Hasil analisis memiliki persyaratan terhadap unidimensionalitas minimal 20% dapat terpenuhi; apabila lebih 40% artinya lebih bagus; dan lebih 60% artinya istimewa.

Indeks Kesukaran Aitem (b)

Parameter kesukaran aitem merupakan parameter lokasi yang menunjukkan posisi item characteristic curve (ICC) dalam skala kemampuan. Semakin mudah aitem

atau semakin rendah kemampuan letak KKA (Karakteristik Kurva Aitem) semakin ke kiri dan begitupula sebaliknya (Hambleton dkk, 1991).

Pengelompokan tingkat kesukaran butir aitem dapat dilakukan dengan menggunakan nilai standar deviasi (SD) (Sumintono & Widhiarso, 2015). Menurut Sumintono dan Widhiarso (2015) lebih besar dari +1SD adalah kelompok soal yang tergolong sangat sukar; 0.0 logit +1SD adalah kelompok soal yang tergolong soal sukar; 0.0 logit -1SD adalah kelompok soal yang tergolong soal mudah; dan lebih kecil dari +1SD adalah kelompok soal yang tergolong sangat mudah.

Tabel 1. Dasar Penetapan Aitem Revisi

No	IKA-IRT	Outfit		Pt. Measure		DIF-Gender	Kesimpulan
		MNSQ	ZSTD	Corr.	Exp.		
157	-0.08	0.97	-0.3	g. 0.50	0.45	Tidak	Tetap
158	-0.37	0.88	-1.0	e. 0.46	0.43	Tidak	Tetap
159	-0.27	0.97	-0.2	h. 0.46	0.43	Tidak	Tetap
160	-1.41	0.79	-0.9	d. 0.39	0.33	DIF Laki-laki	Revisi
161	-0.41	0.92	-0.6	j. 0.44	0.42	Tidak	Tetap
162	-1.14	0.76	-1.3	b. 0.43	0.36	Tidak	Tetap
163	-0.43	0.82	-1.5	c. 0.48	0.42	Tidak	Tetap
164	-1.02	0.86	-0.8	f. 0.41	0.37	Tidak	Tetap
165	0.03	1.23	2.2	B. 0.39	0.46	Tidak	Tetap
166	-0.10	1.19	1.7	E. 0.39	0.45	DIF Laki-laki	Revisi
167	0.25	1.24	2.5	A. 0.39	0.48	Tidak	Revisi
168	0.92	0.98	-0.2	l. 0.52	0.52	Tidak	Tetap
169	0.86	0.82	-2.2	a. 0.59	0.51	Tidak	Tetap
170	0.31	0.99	-0.1	H. 0.46	0.48	Tidak	Tetap
171	-0.29	0.91	-0.7	i. 0.45	0.43	Tidak	Tetap
172	-0.12	1.20	1.8	D. 0.38	0.45	Tidak	Tetap
173	0.69	1.05	0.6	G. 0.47	0.50	Tidak	Tetap
174	0.92	1.22	2.3	C. 0.45	0.52	Tidak	Tetap
175	1.17	0.99	-0.1	J. 0.55	0.53	Tidak	Tetap
176	0.49	1.08	1.0	F. 0.47	0.49	Tidak	Tetap

Differential Item Functioning (DIF)

Analisis aitem dalam IRT dapat dilakukan menggunakan DIF untuk mendeteksi adanya bias respon yang disebabkan perbedaan karakteristik antara satu kelompok dengan kelompok yang lain. Bias dalam aitem dapat diketahui berdasarkan nilai yang berada dibawah 5% (0,05) pada tabel PROB. yang terdapat pada program winstep (Sumintono & Widhiarso, 2015).

Aitem Fit Order

Aitem fit order digunakan untuk melihat dan memberikan informasi sejauhmana suatu aitem dapat dikatakan fit atau misfit dalam mengukur suatu konstruk yang dilihat dari item fit order. Aitem dapat dikatakan fit atau misfit dapat diketahui melalui nilai INFIT dari MNSQ dengan kriteria: nilai MNSQ (Outfit Mean Square) yang diterima ($0.5 < MNSQ < 1,5$); nilai ZSTD (Outfit Z-Standard) yang di-

terima ($-2.0 < ZSTD < +2.0$) dan nilai Pt Mean Corr (Point Measure Correlation) dengan nilai $0.4 < Pt\ Measure\ Corr < 0.85$ (Sumintono & Widiarso, 2013).

Metode

Sampel Penelitian

Sampel penelitian adalah mahasiswa UIN Suska Riau tahun pertama (T.A. 2016/2017) yang mengisi tes IST sebanyak 417 orang, yang terbagi ke dalam dua tahapan penelitian. Pada tahap pertama berjumlah 293 orang dan tahap kedua berjumlah 124 orang.

Tahapan Penelitian

Pada penelitian ini terbagi atas dua tahapan. Tahap pertama ialah melakukan analisis terhadap subtes ME dan mereduksi aitem-aitem subtes ME yang termasuk dalam kategori tidak baik berdasarkan karakteristik psikometri. Aitem-aitem yang termasuk kategori tidak baik akan direduksi dan dilakukan revisi. Aitem yang telah direvisi akan dilakukan uji coba kembali dan akan dianalisis ulang.

Teknik Analisis

Analisis data yang digunakan ialah analisis deskriptif menggunakan pemodelan rasch untuk analisis dan mereduksi aitem yang termasuk kategori tidak baik dalam batasan psikometri dengan bantuan program winstep. Pengujian yang dilakukan meliputi: uji unidimensionalitas, reliabilitas, indeks kesukaran aitem, item fit order, dan DIF (Differential Item Functioning) berdasarkan jenis kelamin.

Hasil

Berdasarkan analisis yang telah dilakukan pada subtes Merkaufgaben (ME), maka dapat disimpulkan terdapat 1 aitem yang tidak fit, yaitu aitem nomor 167. Hal ini dikarenakan aitem nomor 167 tidak memenuhi kriteria nilai ZSTD dan Pt. Mean Corr. Selain itu, terdapat dua aitem yang terjangkau DIF (Differential Item Functioning) terhadap jenis kelamin, aitem nomor 160 dan 166. Revisi yang dilakukan ialah dengan mengubah kata-kata hafalan pada subtes ME. Penetapan aitem yang direvisi dapat dilihat pada tabel 1 berikut:

Tabel 2. Uji Unidimensionalitas ME Versi Asli dan Revisi

		Asli		Revisi	
		Empirical	Modeled	Empirical	Modeled
Total raw variance in observations	30.1	100%	100%	100%	100%
Raw variance explained by measures	10.1	33.6%	25.4%	25.4%	25.4%
Raw variance explained by persons	3.7	12.2%	11.9%	11.9%	11.9%
Raw variance explained by items	6.5	21.4%	13.5%	13.5%	13.5%
Raw unexplained variance (total)	20.0	66.4%	74.6%	74.6%	74.6%
Unexplained variance in 1st contrast	2.0	6.5%		6.4%	
Unexplained variance in 2nd contrast	1.9	6.3%		5.7%	
Unexplained variance in 3rd contrast	1.7	5.8%		5.6%	
Unexplained variance in 4th contrast	1.5	5.1%		5.0%	
Unexplained variance in 5th contrast	1.5	5.0%		4.7%	

Unidimensionalitas

Pada tabel 2 terlihat hasil pengukuran raw variance data adalah sebesar 33.6%. Nilai ini mendekati dengan nilai ekpektasinya, yaitu sebesar 33.1%. Hal ini menunjukkan bahwa persyaratan unidimensionalitas sebe-

sar 20% dapat terpenuhi (Sumintono & Widiarso, 2015). Artinya, subtes ME versi revisi mampu menunjukkan bahwa secara umum, responden cenderung mampu untuk menjawab aitem pada subtes ME. Hal ini dikarenakan subjek berasal dari mengukur aspek

yang hendak diukur yaitu daya ingat. Pada tabel 2 dapat dilihat bahwa ME versi revisi mengalami peningkatan nilai raw variance data dari sebelumnya yaitu 25.4% (tabel 2) menjadi 33.6%. Artinya ME versi revisi memberi pengaruh dalam pengukuran konstruk memori pada individu.

Reliabilitas

Hasil analisis reliabilitas subtes ME versi asli diperoleh informasi bahwa jumlah data yang diberikan kepada 293 responden dengan 20 aitem subtes ME adalah sebanyak 5580 data. Hasil analisis yang dilakukan memuat dua buah output, yaitu output untuk person dan output untuk item.

Hasil analisis diperoleh rerata nilai measure pada tabel person adalah 0.52 logit ($\mu > 0.00$). Hal setting demografis yang beragam, baik usia, jenjang pendidikan, dan lembaga pendidikannya.

Nilai cronbach alpha (KR-20) pada analisis ME yang dilakukan menjelaskan nilai reliabilitas, yaitu interaksi antara person dan item secara keseluruhan. Nilai $\alpha = 0.81$ untuk

ME versi asli dan mengalami perubahan nilai pada ME versi revisi dengan $\alpha = 0.77$. Sesuai dengan kriteria yang digunakan apabila nilai α sebesar 0.7 – 0.8 termasuk dalam kategori bagus (Sumintono & Widhiarso, 2015). Artinya subtes ME masih tergolong reliabel dalam mengukur daya ingat individu.

Indeks Kesukaran Aitem

Indeks kesukaran aitem pada subtes ME. Pengelompokan tingkat kesukaran butir aitem dapat dilakukan dengan menggunakan nilai standar deviasi (SD) (Sumintono & Widhiarso, 2015). Nilai logit yang tinggi menunjukkan tingkat kesukaran soal yang tinggi (Sumintono & Widhiarso, 2015). Menurut Sumintono dan Widhiarso (2015) lebih besar dari +1SD adalah kelompok soal yang tergolong sangat sukar; 0.0 logit +1SD adalah kelompok soal yang tergolong soal sukar; 0.0 logit -1SD adalah kelompok soal yang tergolong soal mudah; dan lebih kecil dari +1SD adalah kelompok soal yang tergolong sangat mudah. seperti tabel berikut:

Tabel 3. Kriteria Indeks Kesukaran Aitem

No.	Kriteria (logit)	Ket
1.	> 0.69	Sangat sukar
2.	0.0 – 0.69	Sukar
3.	0.0 – (-0.69)	Mudah
4.	< (-0.69)	Sangat mudah

Ket : Standar deviasi : 0.69

Mengacu pada tabel 4, indeks kesukaran aitem subtes ME bergerak dari rentang -1.41 logit hingga 1.17 logit. Nilai logit yang tinggi menunjukkan tingkat kesukaran soal yang tinggi (Sumintono & Widhiarso, 2015). Maka dapat disimpulkan bahwa subtes ME terbagi dalam empat kelompok kesukaran

aitem, yaitu 4 soal termasuk kategori sangat sukar, 1 soal kategori sukar, 12 soal kategori mudah, dan 3 soal kategori sangat mudah.

Pengelompokan aitem berdasarkan kesukaran aitem subtes ME dapat dilihat sebagai berikut:

Tabel 4. Indeks Kesukaran Aitem Versi

Aitem	B	
175	1.17 logit	Sangat sukar
168	0.92 logit	Sangat sukar
174	0.92 logit	Sangat sukar
169	0.86 logit	Sangat sukar
173	0.69 logit	Sukar
176	0.49 logit	Mudah
170	0.31 logit	Mudah
167	0.25 logit	Mudah
165	0.03 logit	Mudah
157	-0.08 logit	Mudah
166	-0.10 logit	Mudah
172	-0.12 logit	Mudah
159	-0.27 logit	Mudah
171	-0.29 logit	Mudah
158	-0.37 logit	Mudah
161	-0.41 logit	Mudah
163	-0.43 logit	Mudah
164	-1.02 logit	Sangat mudah
162	-1.14 logit	Sangat mudah
160	-1.41 logit	Sangat mudah

Differential Item Functioning (DIF)

Bias aitem dalam pengukuran ini dilihat berdasarkan jenis kelamin. Deteksi bias aitem dilakukan untuk melihat apakah terdapat perbedaan probabilitas menjawab setiap butir soal merkaufgaben (ME) pada kelompok responden berjenis kelamin laki-laki dan perempuan. Analisis model Rasch menampilkan deteksi bias aitem dalam keberfungsian aitem diferensial (Differential Item Functioning atau DIF).

Pada aitem 160, nilai prob. sebesar 0.8302 dan aitem no 166 memiliki nilai prob. (p) sebesar 0.5346. Kriteria bias suatu aitem dapat dilihat melalui nilai prob. dengan nilai <5% atau <0.05 (Sumintono & Widhiarso, 2013). Pada tabel 5, aitem nomor 160 dan 166 mengalami perubahan nilai. Artinya secara keseluruhan aitem-aitem subtes ME versi revisi bebas dari bias jenis kelamin.

Tabel 5. Perbandingan Analisis DIF

No.	Person Classes	Prob.	
		Asli	Revisi
157	2	0.4741	0.4870
158	2	0.2028	0.8174
159	2	0.2674	0.4269
160	2	0.0179	0.8302
161	2	0.117	0.7018
162	2	0.3846	0.4616
163	2	0.5367	0.4764
164	2	0.1130	0.0670
165	2	1.000	0.1348
166	2	0.0336	0.5346
167	2	0.0971	0.9541
168	2	0.7597	0.2054
169	2	0.1150	0.2894
170	2	0.0511	0.9541
171	2	0.3099	1.000
172	2	0.7234	0.2496
173	2	0.1438	0.1663
174	2	0.0772	0.9259
175	2	0.5845	1.000
176	2	0.6431	0.1627

Aitem Fit Order

Aitem Fit Order merupakan analisis yang berkaitan dengan sejauh mana keber-

fungsian aitem pada subtes ME mampu mengukur daya ingat individu (Sumintono & Widhiarso, 2015).

Tabel 6. Aitem Fit Order

No.	Outfit		Pt. Measure	
	MNSQ	ZSTD	Cor.	Exp.
157	0.92	-0.5	d. 0.48	0.44
158	1.05	0.4	l. 0.41	0.44
159	0.92	-0.2	e. 0.43	0.40
160	1.17	0.9	F. 0.44	0.42
161	1.07	0.4	G. 0.40	0.42
162	0.75	-0.9	a. 0.49	0.39
163	1.27	1.4	C. 0.31	0.43
164	0.75	-0.8	g. 0.43	0.38
165	0.93	-0.3	J. 0.43	0.44
166	0.80	-1.4	b. 0.53	0.45
167	0.91	-0.6	i. 0.48	0.45
168	1.17	0.8	E. 0.39	0.45
169	1.25	1.5	D. 0.43	0.46
170	0.97	-0.2	h. 0.47	0.45
171	0.85	-0.6	f. 0.45	0.41
172	1.00	0.1	H. 0.41	0.44
173	0.90	-0.6	j. 0.47	0.45
174	3.05	1.9	A. -0.04	0.26
175	0.90	-0.6	c. 0.51	0.46
176	1.33	0.8	B. 0.27	0.36

Hasil analisis subtes ME versi asli dan revisi mengalami perubahan nilai dapat dilihat pada tabel 6. Mengacu pada tabel 6, maka dapat ditarik kesimpulan bahwa aitem nomor 167 memiliki nilai MNSQ sebesar 0.91 dengan kriteria $0.5 < \text{MNSQ} < 1.5$, ZSTD sebesar -0.6 dengan kriteria $-2 < \text{ZSTD} < +2$, dan nilai Pt. Mean Corr sebesar 0.48 dengan kriteria $0.4 < \text{Pt. Mean Corr} < 0.85$. Aitem 167 versi revisi sesuai dengan kriteria yang digunakan. Hasil analisis ini diasumsikan karena pemilihan dan pengubahan kata yang dilakukan oleh peneliti sesuai dengan konteks dan tidak terlalu familiar oleh subjek. Sehingga nilai yang diperoleh sesuai dengan kriteria.

Pembahasan

Secara umum, aitem-aitem subtes Merkaufgaben (ME) memiliki karakteristik psikometri yang baik. Hal ini dapat dilihat dari nilai reliabilitas aitem subtes ME yang memiliki tingkat reliabilitas yang bagus dalam mengukur daya ingat individu. Pendapat ini sejalan dengan penelitian yang dilakukan oleh Agung dan Fitri (2016) bahwa subtes ME merupakan salah satu subtes yang memiliki nilai reliabilitas bagus. Selain itu, analisis juga menunjukkan nilai unidimensionalitas yang sudah terpenuhi.

Subtes ME memiliki bentuk soal dan jawaban yang sama persis. Dengan demikian, mengindikasikan aitem-aitem subtes ME memiliki karakteristik psikometri yang baik. Hal ini didukung oleh penelitian yang dilakukan oleh Bawono (2008) yang menyatakan bahwa secara keseluruhan aitem pada subtes ME memiliki diskriminasi yang baik, sehingga semua aitem mampu mengukur konstruk yang diukur. Meskipun demikian, dari dua puluh aitem yang terdapat pada subtes ME, terdapat satu aitem yang tergolong misfit (aitem 167) dan dua aitem yang termasuk kategori bias terhadap jenis kelamin (aitem 160 dan 166), karena memiliki nilai dibawah kriteria yang ditentukan (Sumintono & Widhiarso, 2015).

Aitem nomor 160 versi asli merupakan aitem yang termasuk pada kategori bias terhadap jenis kelamin. Aitem ini lebih mengun-

tungkan apabila dijawab oleh subjek berjenis kelamin perempuan. Namun, setelah dilakukan revisi terjadi peningkatan nilai prob. yaitu 0.0179 menjadi 0.8302 (lihat tabel 5). Aitem nomor 160 versi asli menggunakan hafalan kata yang termasuk kategori bunga yang tergolong familiar bagi subjek berjenis kelamin perempuan. Hal ini tentunya mengindikasikan terjadinya bias, sehingga dengan melakukan revisi pada hafalan kata tersebut maka dapat menghindari bias pada jenis kelamin. Berdasarkan uraian diatas, maka aitem nomor 160 versi revisi layak digunakan dan tidak diperlukan peninjauan ulang.

Aitem nomor 166 versi asli juga termasuk dalam kategori bias terhadap jenis kelamin laki-laki dan akan menguntungkan apabila dijawab oleh perempuan. Namun, setelah dilakukan revisi, terjadi perubahan nilai prob., yaitu 0.0336 menjadi 0.5346 (lihat tabel 5). Penggunaan jumlah sampel yang dilakukan dalam try out juga menentukan bias atau tidaknya suatu aitem. Sehingga, dari beberapa uraian tersebut maka aitem nomor 166 tidak perlu dilakukan peninjauan ulang.

Aitem yang direvisi berdasarkan kriteria aitem fit order ialah aitem nomor 167. Hal ini dilakukan karena aitem nomor 167 memiliki nilai ZSTD dan Pt. Mean Corr yang tidak sesuai kriteria yang telah ditentukan (lihat tabel 6). Aitem nomor 167 versi revisi memiliki nilai yang sesuai dengan kriteria yang telah ditentukan. Nilai MNSQ sebesar 0.91 dengan kriteria $0.5 < \text{MNSQ} < 1.5$, nilai ZSTD sebesar -0.6 dengan kriteria $-2 < \text{ZSTD} < +2$, dan nilai Pt. Mean Corr sebesar 0.48 dengan kriteria $0.4 < \text{Pt. Mean Corr} < 0.85$. Hal ini menunjukkan perubahan nilai setelah dilakukan revisi. Berdasarkan hal-hal diatas, maka dapat dinyatakan bahwa subtes ME versi revisi masih mampu untuk mengukur daya ingat individu.

Berdasarkan hal-hal diatas, maka dapat dinyatakan bahwa subtes ME versi revisi masih mampu untuk mengukur daya ingat individu. Pendapat ini didukung oleh penelitian yang dilakukan oleh Rahmawati (2014) bahwa dari ketiga estimasi parameter psikometri yang dilakukan, subtes ME masih tergolong baik dengan persentase parameter kesukaran 100% baik, parameter diskriminasi

% baik, dan peluang tebakan 100% baik. Hal ini tentunya mengingat bahwa subtes ME memiliki bentuk soal hafalan yang kontekstual. Hal lain yang menjadikan subtes ME versi revisi baik dikarenakan pada dasarnya tes IST merupakan alat tes yang sudah baku dan masih layak untuk digunakan.

Hasil analisa pada penelitian ini didasarkan pada sampel yang sebagian sama dengan sebelumnya dan mencakup usia populasi. Sehingga, hasil penelitian ini hanya terbatas pada usia sampel saja, yaitu 17 sampai 22 tahun. Namun, perlu diingat kembali bahwa permasalahan pada aitem subtes ME versi asli tidak hanya pada bias jenis kelamin dan aitem yang misfit. Mungkin saja masalah sesungguhnya, seperti indeks kesukaran aitem atau daya diskriminasi. Meskipun demikian, seperti yang telah diungkapkan sebelumnya bahwa pembahasan dalam penelitian ini hanya terbatas pada analisis DIF dan aitem fit order.

Kesimpulan

Berdasarkan hasil uji karakteristik psikometri subtes Merkaufgaben (ME) pada Intelligenz Struktur Test (IST) terdapat perubahan nilai karakteristik psikometri sebelum dan sesudah direvisi. Hasil penelitian ini dapat ditarik kesimpulan bahwa; berdasarkan analisis psikometri yang dilakukan, ME versi revisi sudah termasuk kategori bagus. Selain itu, aitem-aitem subtes ME versi revisi memiliki perubahan nilai, baik analisis DIF, unidimensionalitas, indeks kesukaran, maupun aitem fit order.

Berdasarkan hasil penelitian, subtes ME versi revisi yang dilakukan dengan menggunakan model rasch, peneliti telah memperoleh nilai dari karakteristik psikometri yang termasuk dalam kategori baik, sehingga penggunaan subtes ME versi revisi pada IST layak digunakan. Meskipun demikian, subtes ME versi revisi ini perlu lebih dikembangkan dengan memberikan pada usia sampel penelitian yang lebih representatif. Hal ini bertujuan agar subtes ME versi revisi akan lebih bagus dan tingkat validitas dan reliabilitas semakin tinggi, sehingga subtes ME versi revisi ini da-

pat digunakan sebagai alat ukur kemampuan mengingat yang lebih jitu dalam proses pengukuran psikologis.

Daftar Pustaka

- Adiwimarta, S.S., Darmojuwono, S., & Hastrick, E.S. (2011). *Kamus Universal Langenscheidt* (Jerman). Jakarta: Katalis.
- Agung, I.M. & Ahyani R. F. (2016). *Analisis Psikometri dan Standarisasi Norma pada Tes Intelligenz Struktur Test (IST) pada Mahasiswa UIN Sultan Syarif Kasim Riau*. Pekanbaru. (Laporan Penelitian): Lembaga Penelitian dan Pengabdian Masyarakat (LP2M) UIN Suska Riau.
- Azwar, S. (2011). *Tes Intelegensi*. Pustaka Pelajar: Yogyakarta.
- Bawono, B. A. (2008). *Uji Aspek-Aspek Psikometri Subtes Merkaufgaben dari Baterai Intelligenz Struktur Test*. Thesis. Jakarta: Fakultas Psikologi Unika Atma Jaya.
- Bond, T. G., & Fox, C.M. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Science*, Third Edition. New York: Routledge.
- Embretson, S.E. & Reise, S.P. (2000). *Item Response Theory For Psychologists*. New Jersey: Lawrence Associates, Publisher.
- Hambleton. R. K. & Swaminathan. H. (1985). *Item response theory*. Boston: Kluwer Nijhoff Publisher.
- Hambleton, R.K., Swaminathan, H., & Rogers. (1991). *MMSS Fundamental Statistic in Psychology and Item Response Theory (Volume 2)*. California: Sage Publication.
- Rachmawati, F.A & Fitri, A. (2014). *Confirmatory Factor Analysis Tes Intelligensi Kolektip Indonesia Tingkat Menengah (TIKI-M)*. *Jurnal Psikologi Pendidikan dan Perkembangan*, 3, 1.
- Schmidt, S. J. & Zobel, R. (1983). *Empirische Untersuchungen Zu Personallichkeits-Variablen von Literatur Produzenten*. Springer: Sage Publication.

- Suharman. (2005). Psikologi Kognitif. Surabaya: Srikandi.
- Sumintardja, E. & Rismiyati EK. (1997). Konsep Dasar serta Strategi Pemahaman Psikodiagnostik. Jakarta: PT. Charoen Pokphand Indonesia.
- Sumintono, B. & Widhiarso W. (2013). Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial. Cimahi: Trim Komunikata.
- Sumintono, B. & Whidhiarso, W. (2015). Aplikasi Pemodelan Rasch pada Assessment Pendidikan. Cimahi: Trim Komunikata.
- Wiratna, A (1993). Intelligenz Struktur Tes. Surabaya: Locita Mandayagun.